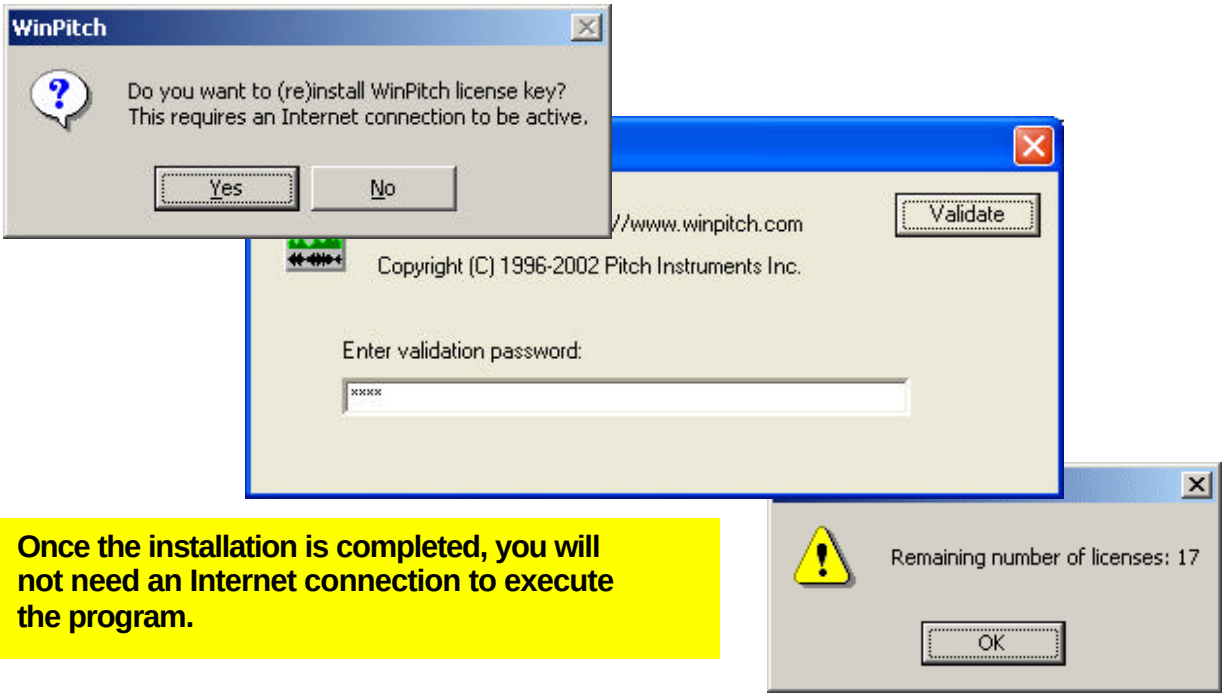
WinPitch - USER MANUAL

1 Installation

The first time you run the WinPitch program, a dialog appears asking for an installation password. If your computer has an active Internet connection at this time, your license password will be activated and the installation will proceed to its final stage. Depending of the number of licenses attached to your password, the same password can be used for multiple installations.



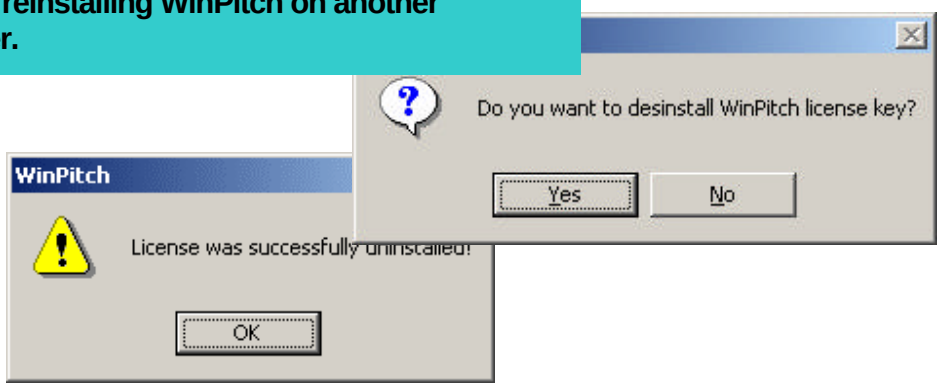
Once the installation is completed, you will not need an Internet connection to execute the program.

Error message if no Internet connection is active on your computer, or if you are connecting through a proxy server.

Your license may also carry a time limit. You can check it by getting the About WinPitch command in the Help menu.

2 Desinstallation

You can also desinstall the program with the same password. Again an active Internet connection is necessary. For desinstallation, start WinPitch Pro keeping your keyboard Control key down. The same license dialog box will appear, prompting you for your password. Once the desinstallation process is completed, you will not be able to run WinPitch on your computer. However, your number of licenses will be incremented by 1, allowing reinstalling WinPitch on another computer.



Computer System Requirements

WinPitch should operate properly on any Pentium III or Pentium IV processor, with at least 20 Mb free memory (RAM) and 6 Mb free disk space. For real-time operation, a Pentium III running at 650 MHz or faster is recommended. The speed of graphic display is also important for real-time spectrogram display. The computer available RAM determines the amount of record and synthesis duration you can use (the program itself uses only 6 Mb).

A color screen with at least 256 colors (16 millions colors recommended) and a minimum resolution of 1025 * 768 is recommended (XGA format).

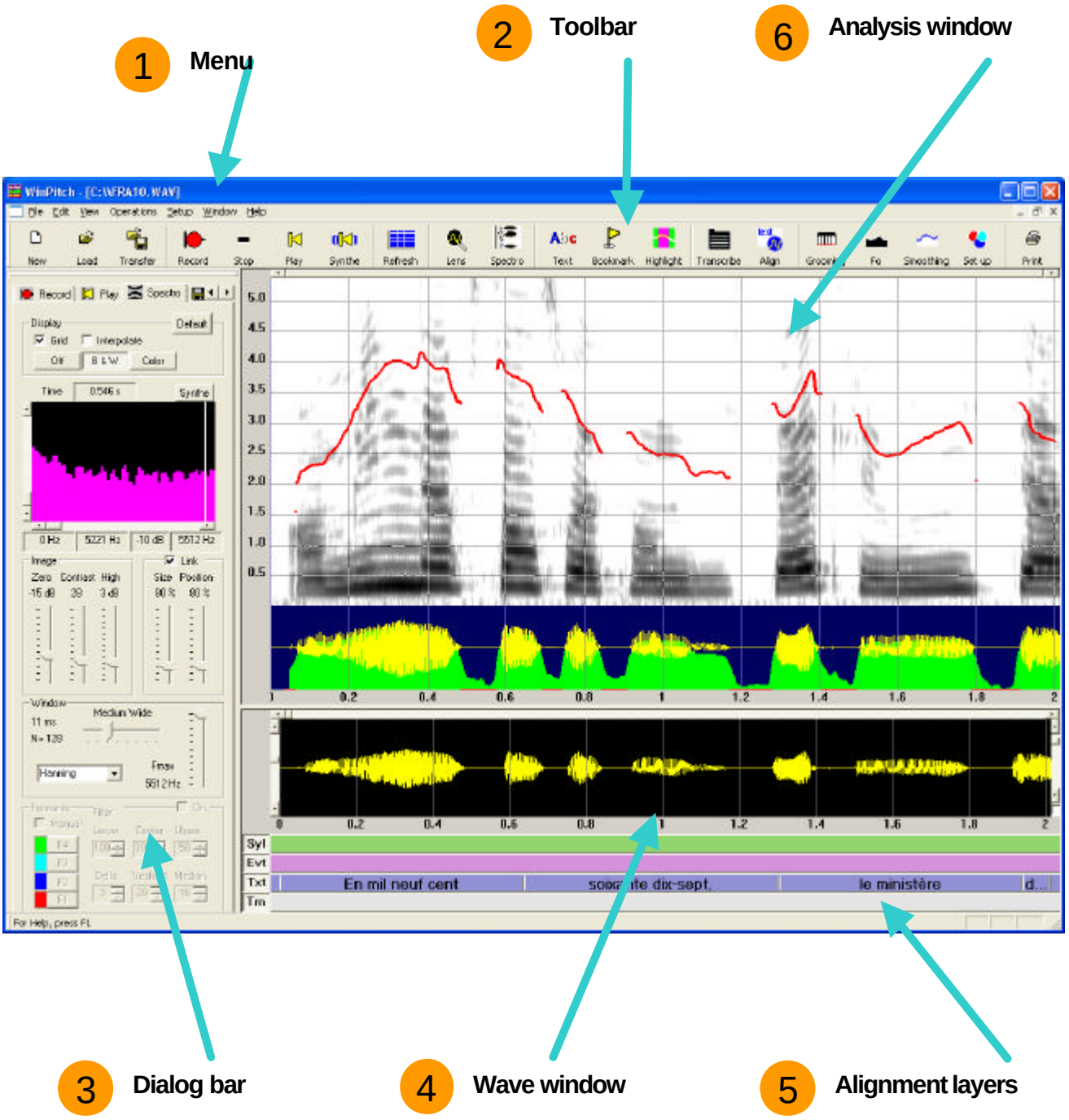
A multimedia sound card for the sound interface (recording and playback) should at least operate at 11025 Hz sampling rate with a mono 16 bit format. (Most current cards operate now on standard 11025 Hz, 22050 Hz and 44100 Hz sampling rates, whereas speech analysis normally uses 22050 or 32000 Hz sampling frequencies). WinPitch supports 8000 Hz, 16000 Hz and 32000 Hz sapling rates.

Current versions of WinPitch run on Windows 2000 and Windows XP operating systems.

2 A first look

WinPitch screen is divided into the following control and display sections:

1. The menu, allowing direct commands by selecting an item in the menu;
2. The tool bar, accessing various commands by clicking on the corresponding icons;
3. The dialog bar, a set of dialog boxes grouping functions and adjustment parameters;
4. A wave window, allowing the selection of section of the signal (sound) to analyze;
5. Alignment layers, displaying aligned text;
6. An analysis window displaying the various analyses performed on the part of signal selected in the wave window.


1 Menu
2 Toolbar
6 Analysis window
3 Dialog bar
4 Wave window
5 Alignment layers
WinPitch - [C:\VFRA10.WAV]
En mil neuf cent
soixante dix-sept
le ministère

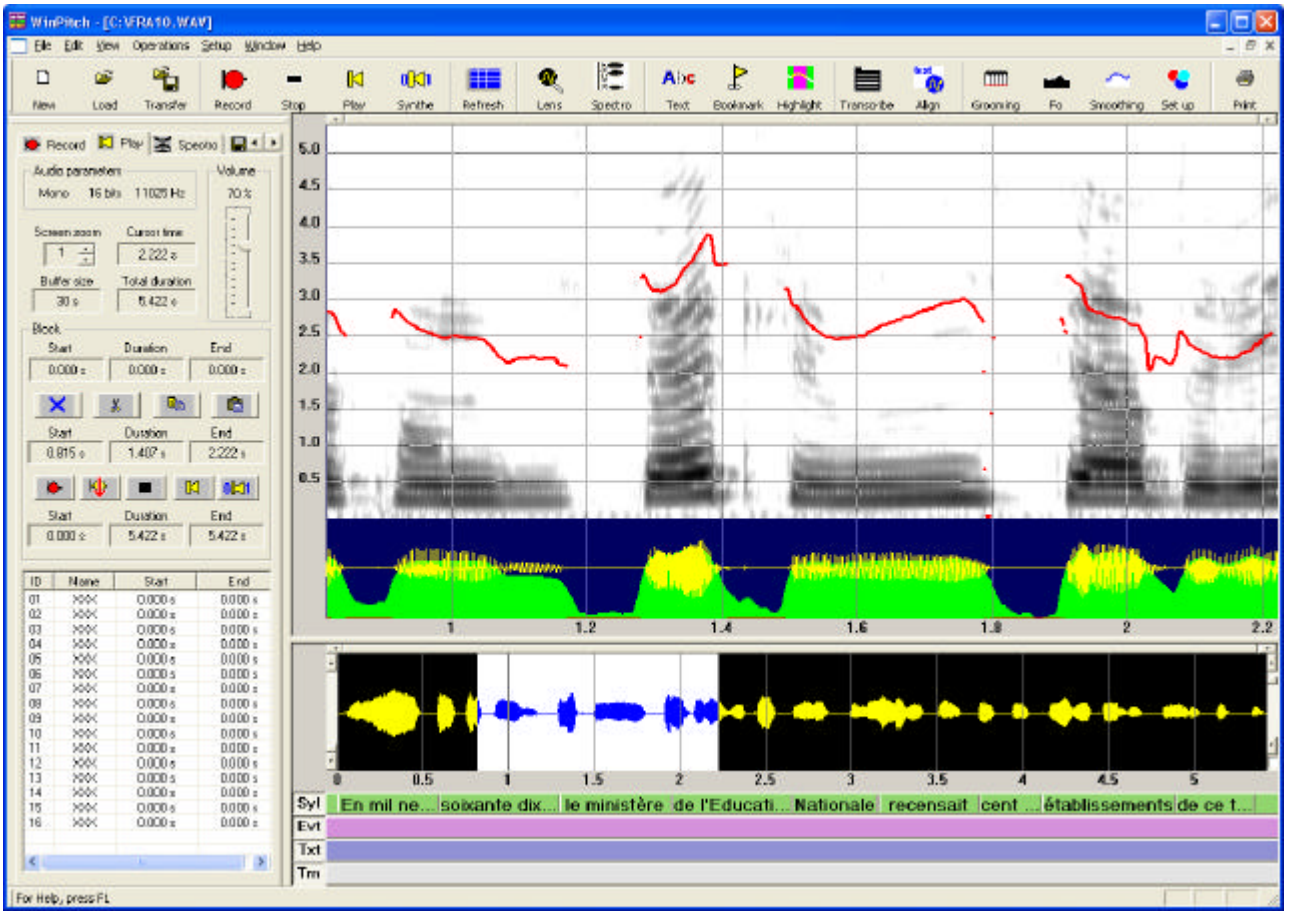
A Wave window zoom and pan

By default, the wave duration displayed in the wave window is equal to the sound duration or 30 seconds, whichever is shorter (this 30 sec. value can be changed in the Setup page).

To change the duration displayed, use the wave zoom vertical slider (a).
To pan the wave window along the signal, use the horizontal wave pan slider (b).
To adjust the wave amplitude displayed in the wave window, use the vertical wave amplitude slider (c).
To invert the wave phase (180 degree shift), click in the wave phase inverter check box (d).

The section of wave displayed in the wave window is re-analyzed and the analysis displayed each time you click on the Refresh icon on the tool bar. The actual display depends on the display options selected (in the illustration, the wave form, the intensity and fundamental frequency curves as well as the wide band spectrogram corresponding to the section of sound signal selected in the wave window are displayed).

B Wave window block


WinPitch - [C:\FRA10.WAV]
File Edit View Operations Setup Window Help
New Load Transfer Record Stop Play Synthe Refresh Lens Spectro Text Bookmark Highlight Transcribe Align Grooming Fo Smoothing Set up Print
Record Play Spectro
Audio parameters
Mono 16 bits 11025 Hz
Volume
70 %
Screen zoom
Cursor time
2.222 s
Buffer size
30 s
Total duration
5.422 s
Block
Start Duration End
0.000 s 0.000 s 0.000 s
0.815 s 1.407 s 2.222 s
0.000 s 5.422 s 5.422 s
ID Name Start End
5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
1 1.2 1.4 1.6 1.8 2 2.2
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
Syl
En mil ne... soixante dix... le ministère de l'Educati... Nationale recensait cent ... établissements de ce t...
Evt
Txt
Tm
For Help, press F1

B Wave window block

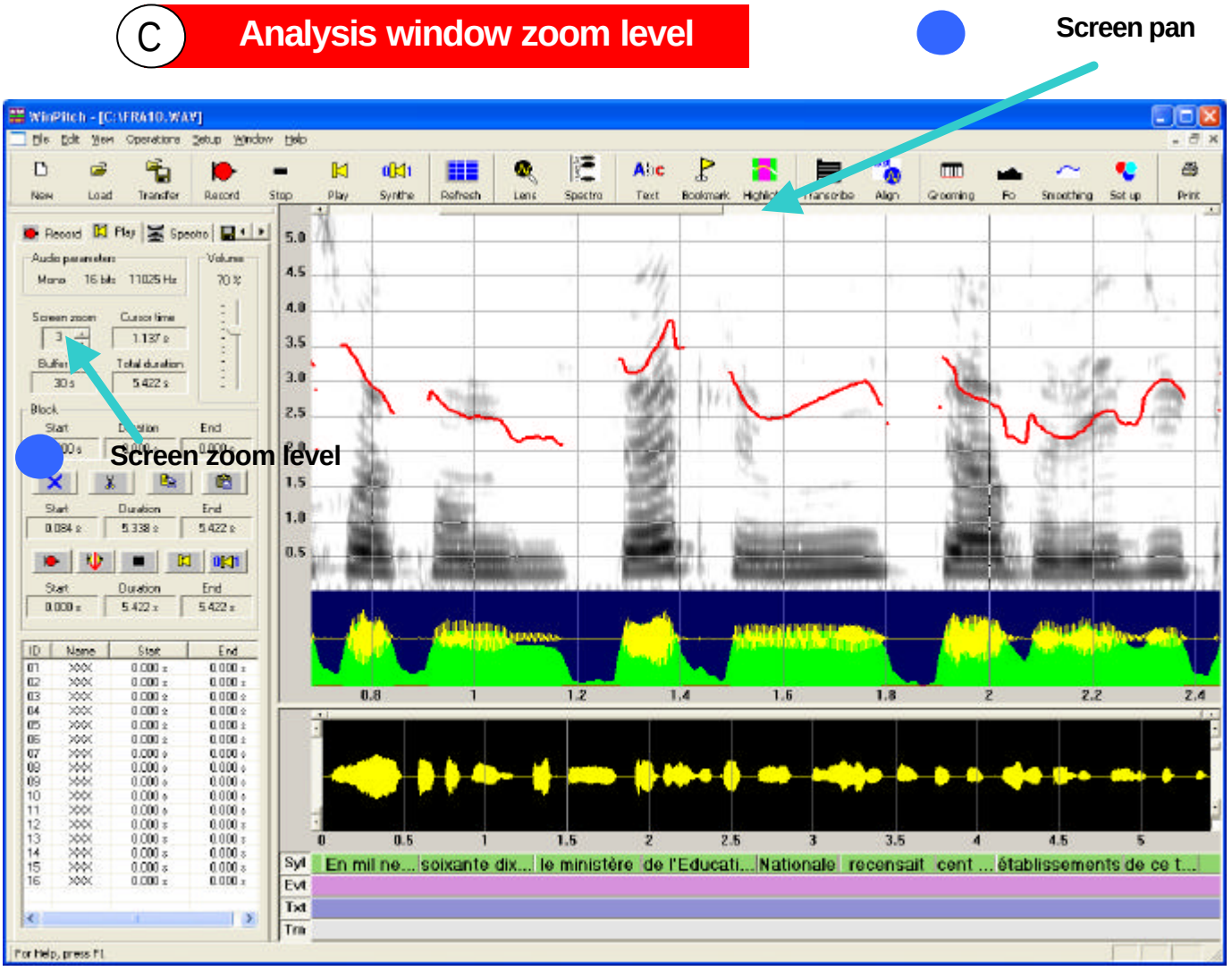
You can also select a section of the wave form by defining a block inside the portion of signal displayed in the wave window.
To define a block, position the mouse cursor on the starting point in time of the block you want to define. Click on the mouse left button and drag the cursor to the ending time position of the block keeping the mouse left button down. Release then the mouse button. The block will be defined (in white on the figure), the analysis window (on top) will be updated to correspond to the signal section defined by the block, and the sound section will be played back (option selectable in the Setup page).

You can move the block by position the cursor about its middle position. The cursor takes then a 4 side arrow shape, indication you can drag the block by keeping the mouse left button down.

You can move the left or right edges of the block by positioning the cursor near the block edge you want to move. Press the mouse left button and drag the cursor to its new position.

For both block changes, the analysis window will be updated and the corresponding sound will be played back when you release the mouse left button.

A block can be erase by clicking outside it anywhere in the wave window, or by pressing the mouse right button.



C Analysis window zoom level

A third zoom function is provided in WinPitch. Concurrently with the A and B signal selection processes, you can define an analysis screen zoom level (box a), which generates a virtual upper window up to 10 times its normal horizontal size. You can then pan this window along the time axis with the horizontal time slider in the analysis window (b), without the need to refresh the analysis.

3 WinPitch aided alignment

Alignment is a basic operation in phonetic research. Previously done manually in a transcription into an either orthographic or phonetic representation, this operation is quite time consuming and difficult to execute automatically through the use of speech recognizers.

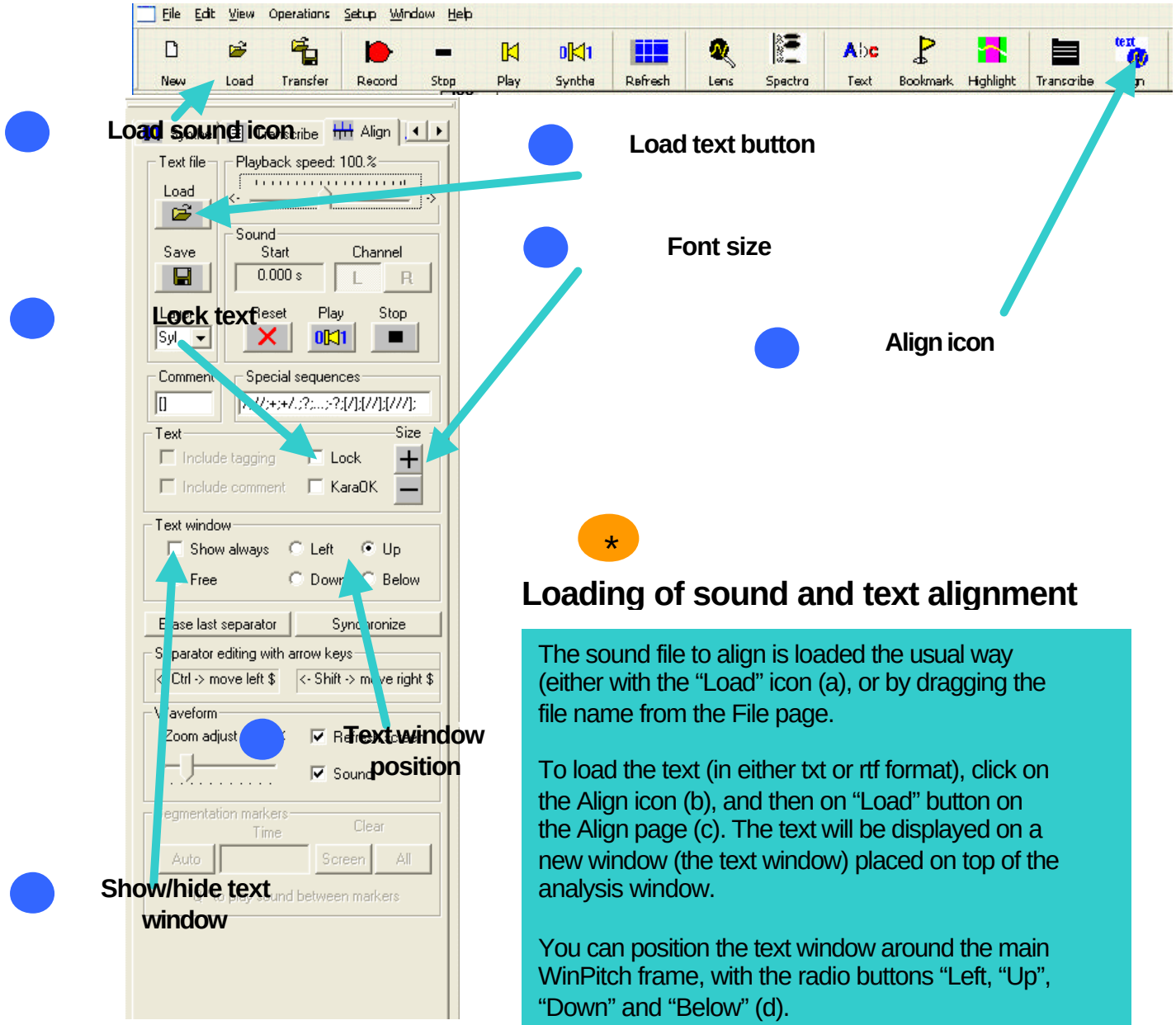
WinPitch has another approach, by allowing the user the select the graphic units as (s)he perceived them. By slowing down the speech playback rate by a selectable factor of 1 to 7, it becomes quite easy to synchronize the operator cursor positioning and mouse clicking with the perceived sound unit, be a sentence, a word or even a syllable. At each mouse click, the ending time of the current unit is automatically defined and entered in the alignment database. At any time of the process, the alignment can be verified and edited by using the karaoke playback function.

Text pre-formatted with the C-ORAL-ROM conventions will be automatically aligned in up to 8 speakers or events levels.

For ease of operation, the playback rate is dynamically adjustable by the user while the alignment process takes place. Inserted unit markers can be easily edited, and text transcription can be entered and edited as well in an alignment session.

The alignment output is generated in XML as well as Excel® formats. Various tools are provided to allow the user to navigate in the alignment database, and obtain the acoustical analysis of aligned units quickly and easily.

A Basic operations



*

Loading of sound and text alignment

The sound file to align is loaded the usual way (either with the “Load” icon (a), or by dragging the file name from the File page.

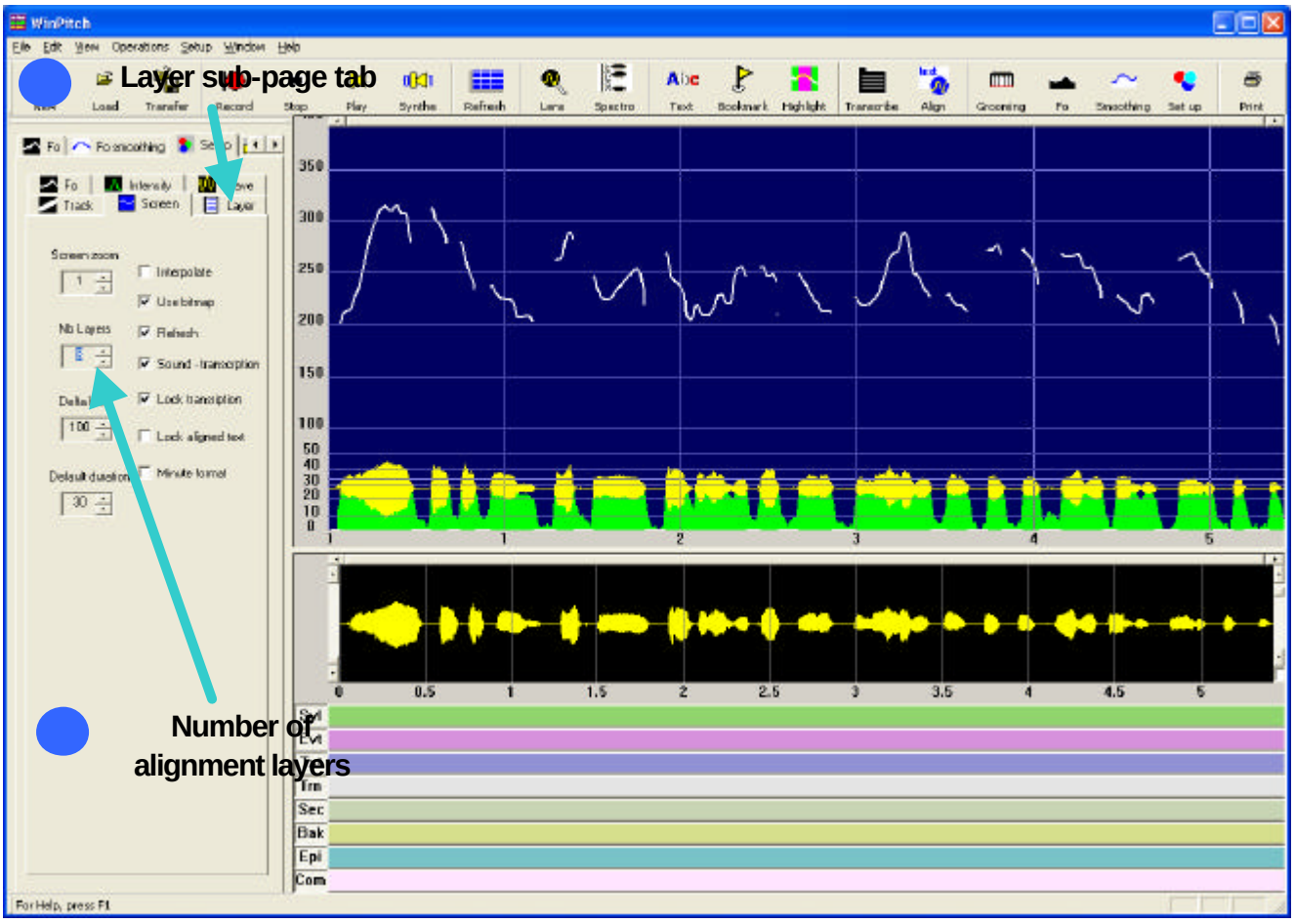
To load the text (in either txt or rtf format), click on the Align icon (b), and then on “Load” button on the Align page (c). The text will be displayed on a new window (the text window) placed on top of the analysis window.

You can position the text window around the main WinPitch frame, with the radio buttons “Left, “Up”, “Down” and “Below” (d).

The text window is hidden when you select another dialog page, but can be set to be displayed permanently with the “Show always” check box (e).

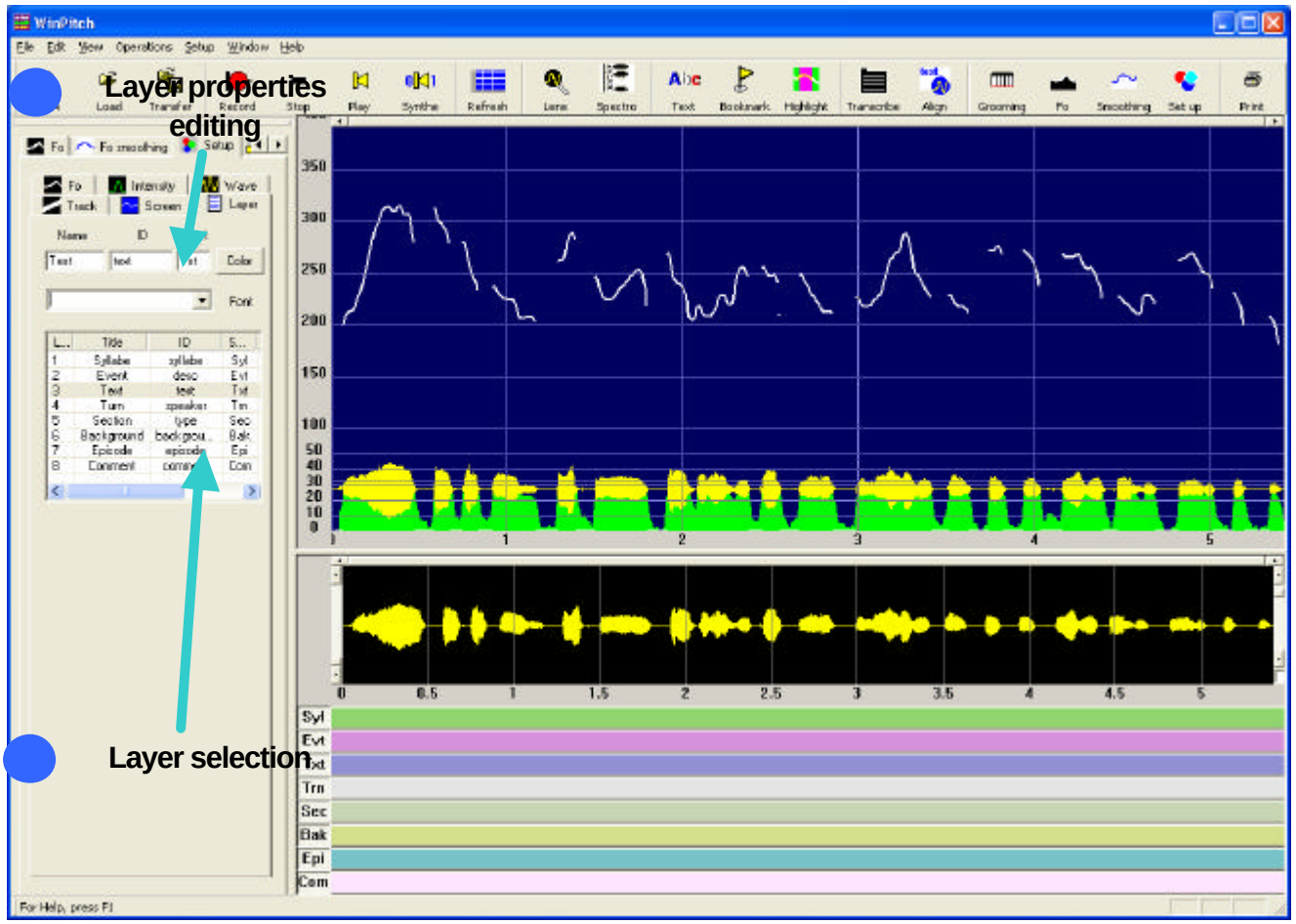
The text font size is adjusted with the “+” and “-“buttons (f), and the text can be protected or be edited with the ‘Lock” check box (g).

B Alignment layers



By default 4 alignment layers are displayed in the wave window. Their number can be adjusted (with a maximum of 8) in the Setup page, Screen sub-page (a).

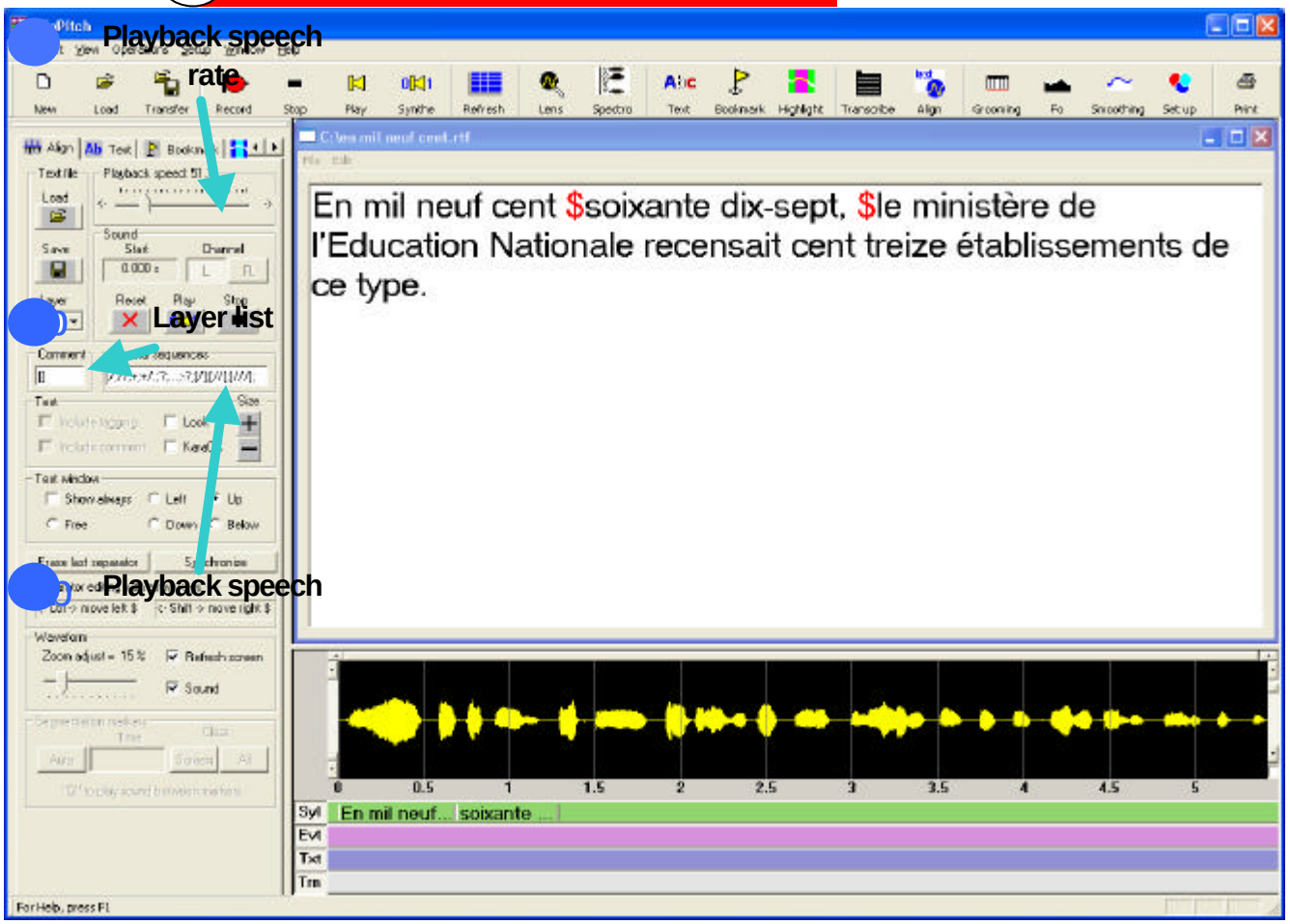
Their names, short names, colors and font can be edited in the Layer sub-page (click on the “Layer” tab (b) to display this sub-page).



To edit layer properties, click on the layer name in the list box (a). You can then edit the layer name and short name (3 characters), and select the color and font which will be used in the wave window (b).

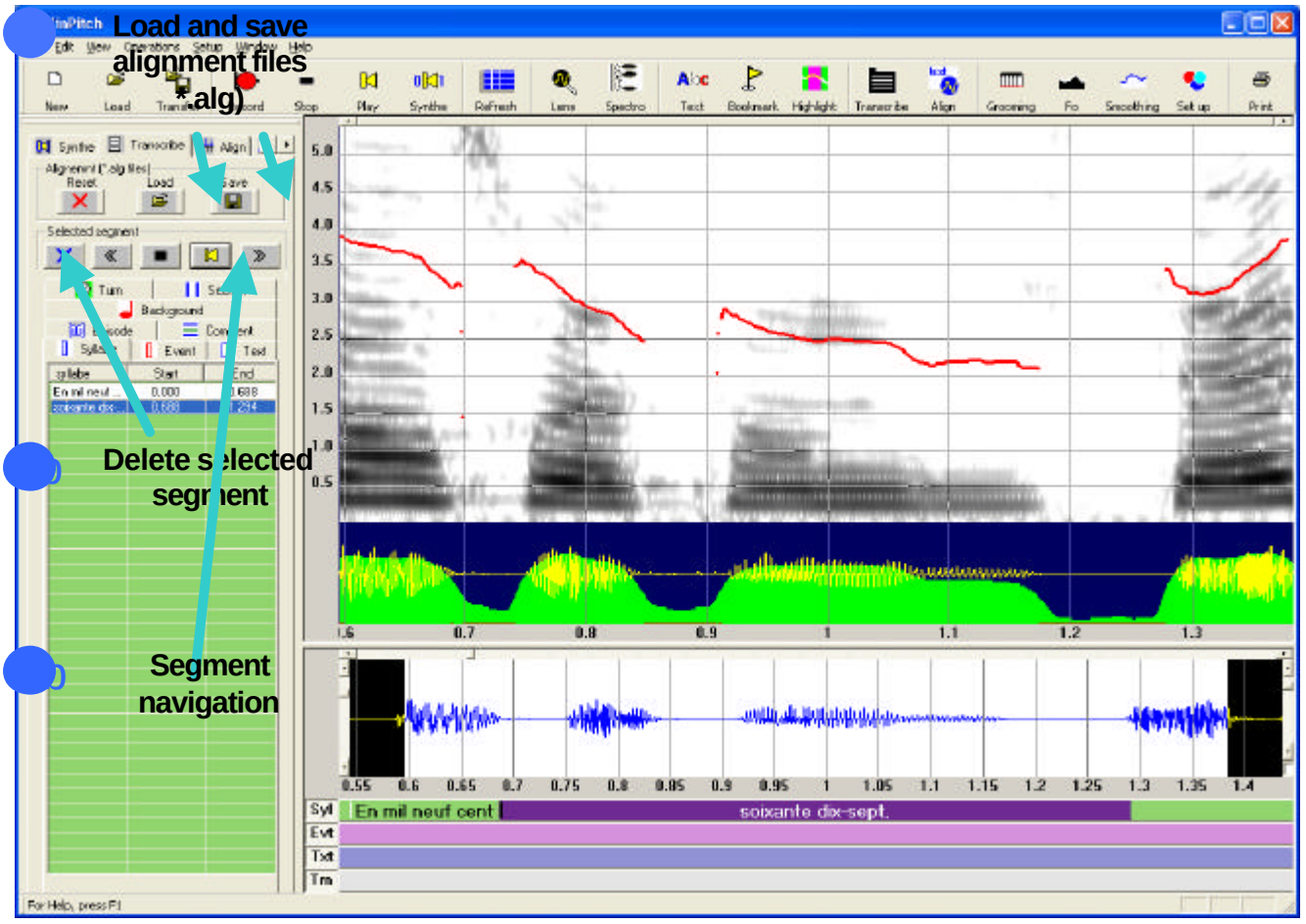
Their names, short names, colors can be edited in the Layer sub-page (click on the “Layer” tab (b) to display this sub-page).

C Semi automatic alignment



To start aligning, you select a comfortable playback speed (a), and you click on the play icon (b). The playback speed can be adjusted on the fly.
For each unit you want to align, you position the mouse cursor on the last graphic part of the text unit (be a sentence, a word or a syllable), and you click on the mouse left button when you are about to hear the END of the corresponding sound.
The unit ending time defined by your mouse click will be automatically entered in the alignment layer selected in the Layer list (c), and a separation sign (a red “$” here) will be inserted in the text.

D Transcribe page

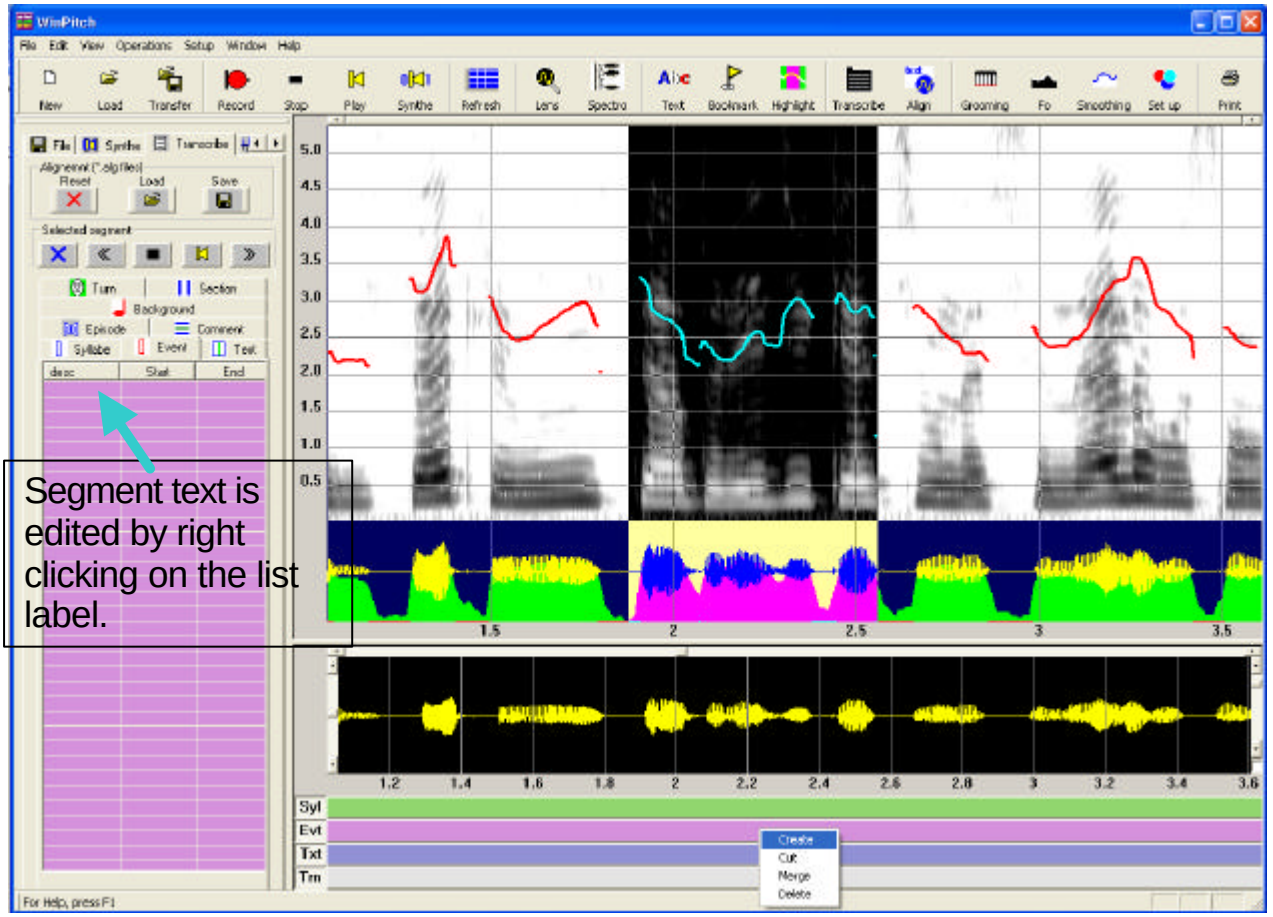


Aligned units are also entered in list tables (one per layer) together with their starting and ending time in seconds. The results are saved in XML format by clicking the “Save” button (a), and can be re-loaded with the “Load” button (a).

Clicking on an aligned unit in one layer list will display the corresponding part of sound in the wave window, and its analysis on the analysis window (unless the text window has been made permanent). Clicking one a unit in the text window or in the wave window will have the same effect.

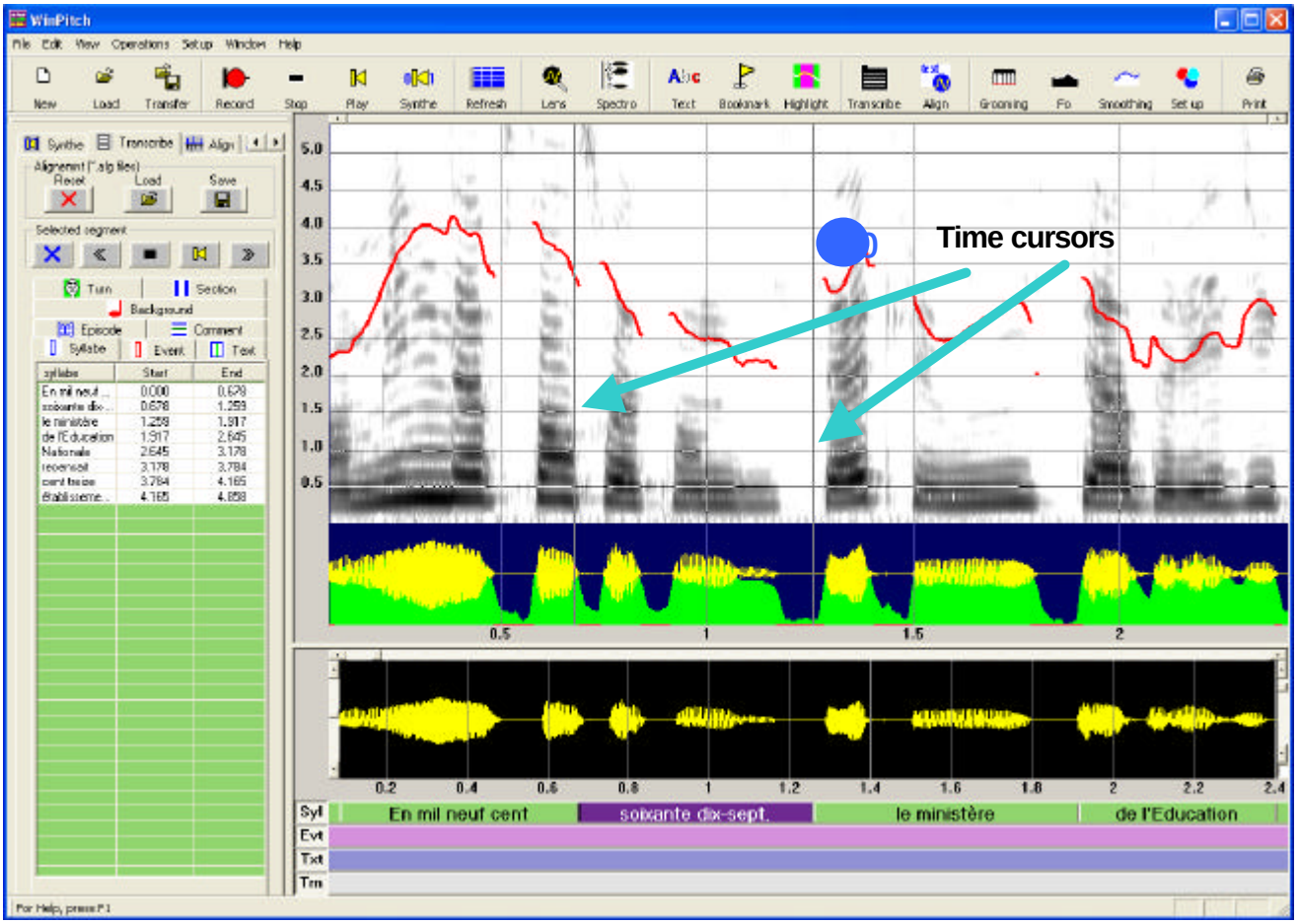
You can also navigate through the list of aligned units with the “Selected segment” buttons (b). A selected segment can be deleted with the (c) button.

E Define direct segments



You can define a segment by positioning the cursor on an alignment layer and RIGHT clicking the mouse button. You then select the desired option: Create a segment, Cut a segment in half, Merge a segment with an adjacent segment or delete a selected segment.

F Editing segments



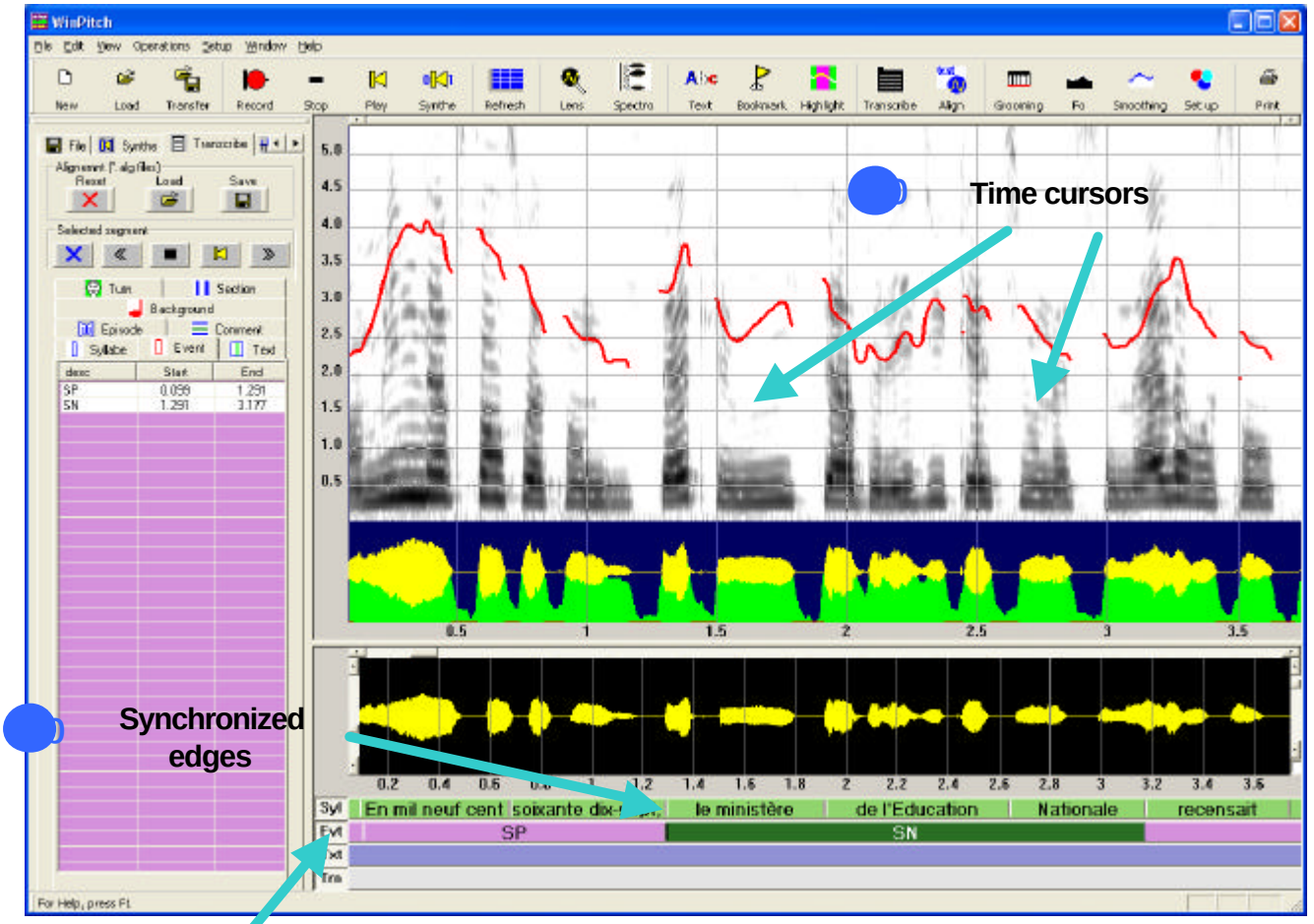
Selected segments can be edited, moved or have their starting and ending time changed.

To select a segment, click on it on the appropriate alignment layer.

Move the cursor on it. Near the edges, the cursor takes a double arrows shape. You can the drag the underlying edge by keeping the mouse left button down. When the cursor is positioned near the middle of the segment, it takes the shape of a quadruple arrow. You can then move the entire segment without changing its size.

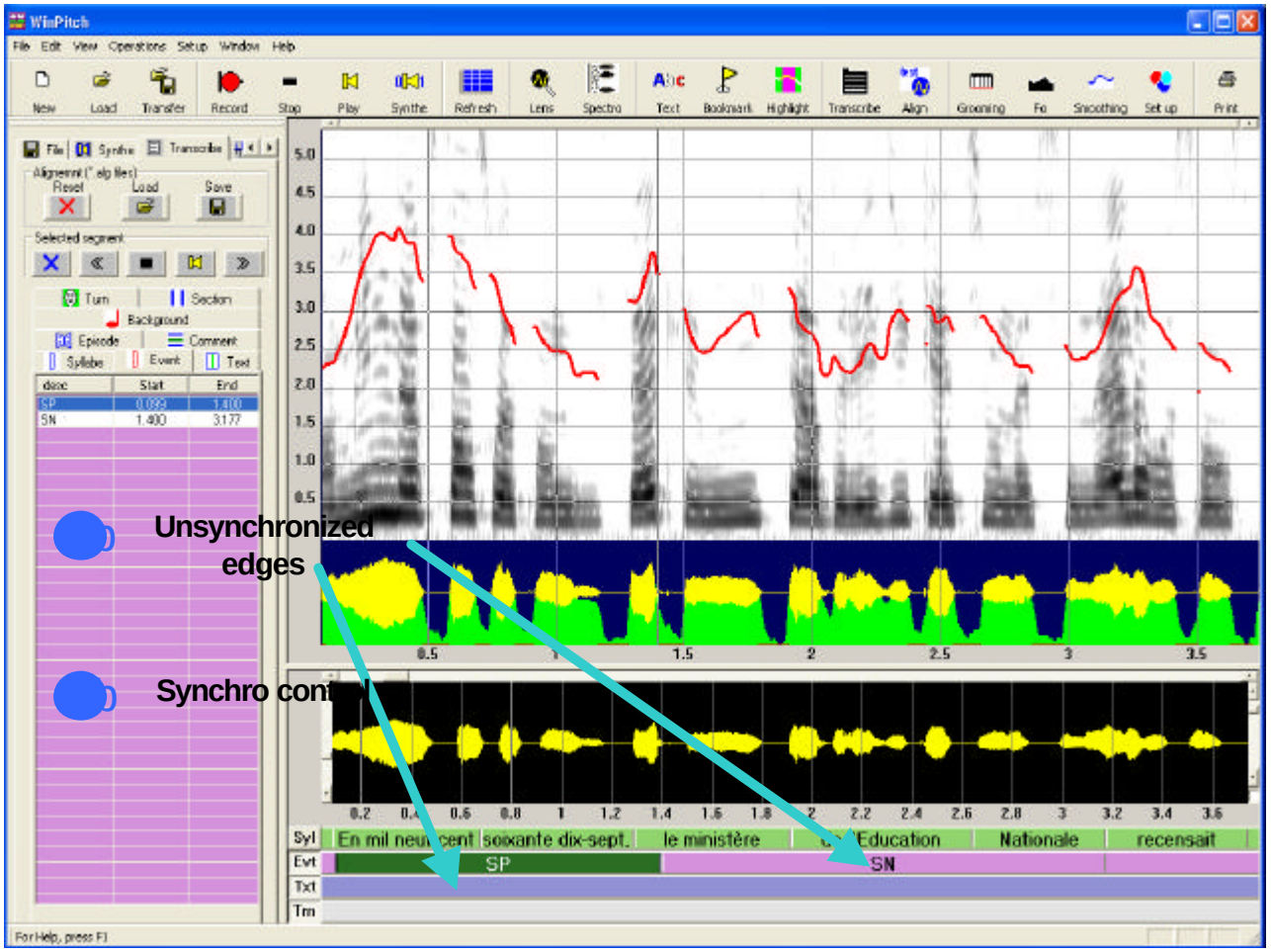
For all the operations, one or two cursors will appear on the analysis window (a), allowing you the achieve a precise alignment if for instance the spectrogram display option is on for the analysis window.

G Layer hierarchy



Changing the edge position of a segment would update segment edges synchronized with it (a).

In the example on the right, moving the edge between SP and SN on layer Evt did move the corresponding edge between "…dix-sept" and "le ministère…".



Layers are synchronized by default.

To turn this feature off, release the layer button on the left by clicking on it (b).

G Output XML format

```
<?xml version="1.0" encoding="iso-8859-1" ?>
<!DOCTYPE Alignment (View Source for full doctype...)>
- <Alignment>
    <TimeStamp Value="Thursday, May 30, 2002 time 00h 10m 04s" />
    <WinPitch Program="Aligner" Version="1.0" />
    <Trans version="1.0" creationDate="Thursday, May 30, 2002 time 00h 10m 04s" audioFilename="C:\\FRA10.WAV" textFilename="C:\en mil
      neuf cent.rtf" />
    <Layer1 Name="Syllabe" ID="syllabe" Short="Syl" Color="RGB(145,213,110)" />
    <Layer2 Name="Event" ID="desc" Short="Evt" Color="RGB(213,145,220)" />
    <Layer3 Name="Text" ID="text" Short="Txt" Color="RGB(145,145,213)" />
    <Layer4 Name="Turn" ID="speaker" Short="Trn" Color="RGB(228,228,228)" />
    <Layer5 Name="Section" ID="type" Short="Sec" Color="RGB(200,213,180)" />
    <Layer6 Name="Background" ID="background" Short="Bak" Color="RGB(213,222,140)" />
    <Layer7 Name="Episode" ID="episode" Short="Epi" Color="RGB(120,195,200)" />
    <Layer8 Name="Comment" ID="comment" Short="Com" Color="RGB(255,228,255)" />
    <UNIT speaker="Syllabe" startTime="0.000" endTime="1.244" Channel="M">En mil neuf cent soixante dix-sept,</UNIT>
    <UNIT speaker="Syllabe" startTime="1.244" endTime="3.137" Channel="M">le ministère de l'Education Nationale</UNIT>
    <UNIT speaker="Syllabe" startTime="3.137" endTime="3.648" Channel="M">recensait</UNIT>
    <UNIT speaker="Syllabe" startTime="3.648" endTime="4.100" Channel="M">cent treize</UNIT>
    <UNIT speaker="Syllabe" startTime="4.100" endTime="4.921" Channel="M">établissements</UNIT>
    <UNIT speaker="Syllabe" startTime="4.921" endTime="5.399" Channel="M">de ce type.</UNIT>
  </Alignment>
```